ENGN 2520 Homework 2
Due Wednesday February 22 by 4pm

Problem 1

Suppose we have two six-sided dice. Dice A has values 1, 2, 3, 4, 5, 6 on its faces. Dice B has value 5 on all faces. Suppose we pick dice A with probability p and dice B with probability 1-p. We toss the selected dice twice and the first time it comes up 5.

(a) What is the probability that we selected dice A given the information we have?

(b) What is the probability that the second toss will come up 5?

(c) What is the expected value of the second toss?

Problem 2

Consider a classification problem in R^D with two classes C_1 and C_2. Suppose we model $p(x|C_i)$ with a Gaussian distribution with class dependent mean $u_i \in \mathbb{R}^D$ and class independent covariance αI where $\alpha \in \mathbb{R}$.

(a) Derive the maximum likelihood estimate for the parameters u_1, u_2, α given training data $\{(x_1, y_1), \dots, (x_n, y_n)\}$.

(b) Suppose we classify an example x by selecting the class with maximum posterior probability $p(C_i|x)$. Show that the decision boundary of the resulting classifier is a hyperplane in R^D.

(c) We say the training data is separated by a hyperplane H if the examples from C_1 is on one side of H and examples from C_2 is on the other side of H.

Describe a training set that can be separated by a hyperplane H but the decision boundary defined by the ML estimation procedure from above does not separate the data.

Problem 3

In this assignment you will implement a naive Bayes classifier to recognize handwritten digits. There are 10 classes corresponding to digits 0 through 9. Each example is a 28x28 binary image represented as a 784 binary vector.

The data for this problem is available on the course website. The training examples are loaded into matrices “train?” where “?” is a digit and the test examples are similarly loaded into matrices "test?". Each matrix has one example per row and the examples can be reshaped into a 28x28 matrix for visualization as follows.

$>$ load('digits');
$>$ A = reshape(train3(43,:))';
$>$ image(A);

Under the naive Bayes model we assume features (pixel values) are independent conditional on the image class. We will use a Bernoulli distribution to model each feature (pixel) of each class. The training and test data are available on the course website.

Let $u(y, i)$ denote the mean of the Bernoulli distribution associated with the i-th feature (pixel) of digit y.

(a) What is the maximum likelihood estimate for the parameters $u(y, i)$ given training data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$?

(b) Use ML estimation to train a model for each digit using the training data from the course website. Make a visualization of the model for each digit by drawing a 28x28 image where the brightness of each pixel specifies the mean of the Bernoulli distribution associated with that pixel.

(c) Use the resulting models to classify the test data. What fraction of the test digits were correctly classified? Compute a 10x10 confusion matrix where entry (i,j) specifies how often digit i was classified as digit j.

You should turn in a writeup that includes a printout of your source code.

Problem 4

In this problem you will consider a geometric interpretation of the classifier from the previous problem. We can think of the input to the classifier as a point in $\mathbb{R}^{768}$. Suppose there are only two digits. Show that the decision boundary for a classifier that picks between two digits using the naive Bayes model from above can be interpreted as a hyperplane in $\mathbb{R}^{768}$. What do the decision regions look like in the case when there are more than 2 digits?